$$\text{Relative frequency of a class} = \frac{\text{Frequency of the class}}{n}$$

% frequency = Relative frequency * 100

n = total sample size

Summarizing data for a quantitative variable: 1. Choose # of non-overlapping classes. 2. Determine class width. Class Width=largest class value minus smallest value divided by # classes. 3. Pick class limits based on class width.

Quantitative value (SAT Scores)	Frequency	Relative Frequency	Percent Frequency
800-999	4	4/14 = .29	29%
1000-1199	8	8/14 = .57	57%
1200-1399	2	2/14 = .14	14%
TOTAL	14	14/14 = 1	100%
Quantitative value (SAT Scores)	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percent Frequency
<= 999	4	4/14 = .29	29%
<= 1199	12	12/14 = .86	86%
<= 1399	14	14/14 = 1	100%

Cumulative # of values <= upper class limit

Sample Mean

$$\overline{x} = \frac{\sum x_i}{n}$$

Weighted Mean

$$\overline{x} = \frac{\sum w_i x_i}{\sum w_i}$$

Location of *p*th Percentile

$$L_p = \frac{p}{100}(n+1)$$

Interquartile range

$$\text{IQR} = Q_3 - Q_1$$

Sample Variance

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$$

Coefficient of Variation

$$\left(\frac{\text{Standard deviation}}{\text{Mean}} \times 100\right)\%$$

z-Score

$$z_i = \frac{x_i - \overline{x}}{s}$$

Sample Covariance

$$s_{xy} = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{n-1}$$

Quartiles

Q1 = 25th percentile

Q3 = 75th percentile

Range

Largest value - smallest value

Standard Deviation

$$s = \sqrt{s^2}$$

Chebyshev

$$1 - \frac{1}{z^2}$$

Five Number Summary

Min
Q1
Median
Q3
Max

Sample Correlation Coefficient

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

(correlation coefficient) where…

r_{xy} = sample correlation coefficient
s_{xy} = sample covariance
s_x = sample standard deviation of x
s_y = sample standard deviation of y

Find location of percentiles:

1. Order the data values ascending.
2. Assign a position number to each value, starting with 1, moving left to right.
3. Find location (L_p) in data set.
4. Find position of value:
   Position of percentile in data set = Small value + decimal part from L_p * (Large value – small value)

Skewness:

For data skewed to the left, the skewness is negative; for data skewed to the right, the skewness is positive. If the data are symmetric, the skewness is zero.

Chebyshev:

1. Calculate z-score for lower and upper values. (z-score must be greater than 1 to be applicable)
2. Calculate Chebyshev formula. Convert to %.

Detecting Outliers:

1. Lower limit = Q1 – 1.5(IQR) ← note the minus
2. Upper limit = Q3 + 1.5(IQR) ← note the plus

Box Plot:

1. Plot Q1, the median, and Q3 using a small vertical line. Draw a box around the lines with Q1 and Q3 on the ends.
2. Calculate upper and lower limits for outliers.
3. Extend whiskers from the box to min and max within limits.
4. Plot any outliers.

Areas Under the Curve for Any Normal Distribution